DOCUMENT RESUME

ED 356 871 PS 021 268

AUTHOR Bezruczko, Nikolaus; Zell, Adrienne
TITLE Relational Data Base System Integrates Early Childhood Assessment Data across Prekindergarten, Kindergarten, and Grade 1. Draft.
PUB DATE Mar 93
NOTE 36p.; Paper presented at the Biennial Meeting of the Society for Research in Child Development (New Orleans, LA, March 25-28, 1993).
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Computer Oriented Programs; *Databases; Data Collection; Early Childhood Education; Early Intervention; *Evaluation Methods; Grade 1; Kindergarten; *Kindergarten Children; *Preschool Children; Primary Education; *Student Evaluation
IDENTIFIERS Illinois

ABSTRACT

The shift away from norm-referenced tests to authentic samples of child performance, such as writing and drawing samples, poses serious problems in organizing and analyzing assessments of child performance. This paper proposes a process utilizing the large-scale storage capacity of centralized mainframe computers and the flexibility and efficiency of table-top computers to conveniently access child observation records, efficiently perform relevant statistical analyses, and reliably produce teacher and parent reports of child performance. The process includes a longitudinal database of performance assessments; an organizational strategy relying on computers and relational databases; and the application of the Rasch measurement model. This paper begins by reviewing trends in child evaluation methods, longitudinal databases, and growth models. The bulk of the paper presents the methods and results of a pilot study applying the proposed process to measure the language behavior of a population of urban disadvantaged children in several early childhood programs. Data sources, including field observations, results of standardized tests, teacher assessments of child performance, and performance samples, are listed; and the structure of the Early Childhood Intervention Database is explained. The paper also explains procedures for population sampling, scale construction and vertical linking, and data analyses. Finally, results of the pilot study are presented and discussed. (AC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED356871

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

DRAFT

Relational Data Base System Integrates Early Childhood Assessment Data Across Prekindergarten, Kindergarten, and Grade 1

by

Nikolaus Bezruczko
Chicago, Illinois

and

Adrienne Zell
University of Chicago

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Nikolaus Bezruczko
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC).

Running head: Integrating Child Assessments

PS 021268

To be presented at the 1993 Meeting of the Society for Research in Child Development, New Orleans.

BEST COPY AVAILABLE

DRAFT

Relational Data Base System Integrates Early Childhood Assessment Data Across Prekindergarten, Kindergarten, and Grade 1

Many contemporary researchers and educators suggest that new ideas in the assessment of school learning now make evaluations based on standardized achievement tests obsolete (Wolf, Bixby, Glenn, & Gardner, 1991). Consequently, local school and statewide evaluation practices quickly are shifting from standardized norm-referenced tests based on multiple choice items to authentic samples of child performance (Herman, 1992; O'Neil, 1992). Writing and drawing samples, student science projects, as well as surveys of teachers and parents are providing information and insight about child mental growth in terms that are more developmentally meaningful than stanines, percentiles, or grade equivalences.

This change in practice, however, presents serious practical and theoretical challenges to educational evaluators and child development researchers. First, performance assessments optimally should occur throughout the year and thus provide continuous feedback to parents and students concerning school learning. School personnel in particular need to access multiple student records on a regular basis. Consequently, performance assessments substantially increase the amount of information that is stored, retrieved, and interpreted for a child.

Second, the tremendous increase in information produced by performance assessments raises questions to evaluators such as, "What aspects of performance are necessary to evaluate student learning?", and to researchers, "How much information is needed to understand growth and development?". Not surprisingly, arbitrary decisions currently are made concerning the information that is reported to teachers and parents.

Yet another problem is the need of parents, teachers, and researchers to compare particular child growth to normative standards. Because the validity, continuity, and generalizability of performance assessments are generally uncertain, they raise issues that were less prevalent with standardized achievement tests (Mehrens, 1992).

In this study, we propose to apply advances in 20th century computer and social science measurement technology to the problems of organizing and analyzing assessments of child performance. The large

scale storage capacity of centralized mainframe computers and the flexibility and efficiency of table top computers are exploited to conveniently access child observation records, efficiently perform relevant statistical analyses, and reliably produce teacher and parent reports of child performance. In this process, local school personnel use specialized mainframe computer data base software to select child information that is transmitted to table top computers in local schools. The table top computers in turn have been equipped with data base software and Rasch scaling technology to statistically link teacher ratings, standardized test scores, and children's performance samples across three levels of school enrollment (i.e., preschool, kindergarten, and grade 1). The results of this process should promote a fundamental goal of performance assessments -- accurate communication between parents and teachers.

Background

The sections below describe trends in program evaluation and developmental research that are promoting the adaptation of computer and measurement test models to the problems of evaluating authentic child performances.

Shift in child evaluation methods. A contemporary trend in school evaluation is the collection of child performance samples (i.e., writing and drawing samples, science projects, and so on). Periodically collected, judges evaluate these samples as products of children's learning. Evaluations based on performance samples have gained wide spread support in universities and public schools despite problems that include:

- computer file management of multiple assessment sources
- duplication and redundancy of assessments
- undefined focus of assessments
- inadequate reliability and unknown construct validity
- vertical equating of assessments across grades
- unknown relations between assessments and normative development.

These problems have several sources. First, because evaluations based on performance samples require corroboration by multiple data

sources, they tremendously increase the volume of data collected of each child. Samples, for example, may be collected of composition writing, journals, exhibits, and interviews, as well as teacher ratings, student and parent surveys, school grades, and standardized test scores. The result is a massive data management problem that increases with each year of child enrollment.

Second, multiple data sources commonly result in duplication and redundancy. Assessments, for example, of students' ability to express ideas in writing based on a collection of writing compositions may be duplicated by an assessment of written expression based on a portfolio of creative writing stories. These assessments then are redundant with assessments of child writing ability based on teacher ratings of child performance.

Third, performance assessments tend to be unsystematic and discontinuous in their surveys of learning. Teachers may differ from year to year in their preference for particular samples, and school policies annually may change in their evaluation focus. Furthermore, performance assessments tend to be determined by immediate schooling concerns rather than long term developmental issues (i.e., an emphasis by some parents to learn a particular kind of writing, become proficient with a computer, or learn some new math method). Consequently, evaluation practices that change from year to year raise serious issues concerning the underlying ability dimensions that can be validly inferred from performance assessments and virtually eliminate the possibility of making normative comparisons of growth. (The shift to performance assessments has not eliminated the need of teachers and parents to identify patterns of typical and exceptional mental development.) This problem of continuity is compounded by the dramatic growth that occurs in children's mental ability from year to year, and especially at younger ages, which requires especially rigorous and precise measurement methods.

Finally, the diligent collection of performance samples and their interpretation does not address the instructional or intervention strategies of particular children to promote growth and development. Children have developmental and pedagogic needs that require an insight into learning mechanisms that may actually become obscured the burden of collecting performance samples.

For the reasons above, as well as others, contemporary evaluations of early childhood learning and intervention programs in particular urgently need an efficient and effective approach to the analysis of performance samples. One that (a) shows relations among authentic assessments, standardized measures, and teacher ratings of performance; (b) is useful for the construction of longitudinal growth continuua such as language or math ability; and (c) provides information that is clear and concise.

Longitudinal data bases. Another influence on child assessment practices are developmental researchers who conduct longitudinal studies that commonly include many variables from multiple sources (i.e., interviews, naturalistic field observations, survey responses, and performance samples.) Consequently, a problem in longitudinal developmental research is the integration of multiple data sources into concise and comprehensible growth models.

Contemporary developmental researchers tend to promote multidimensional structural equation models to analyze data from multiple assessment sources, and many applications of this approach support their use (see Rutter, 1988; Von Eye, 1990). Structural equation models, however, tend to become extremely complex and contemporary researchers have expressed reservations concerning their wide spread use (Biddle & Marlin, 1987; Martin, 1987). Other researchers have encountered intractable practical problems concerning theoretical interpretations and policy implications. Consequently, developmental researchers need a method that simplifies the interpretation, as well as description, of complex functional relationships and key mechanisms that influence child development.

Growth models. Although statistical models are widely applied in child development research (Burchinal & Appelbaum, 1991; Hutteniocher, Haight, Bryk, Seltzer, & Lyons, 1991), the adaptation of mathematical measurement models to growth sequences is less established. Wright (1979) emphasized the role of fundamental measurement for understanding growth phenomena, while Loevinger (1968) provided the conceptual linkage between the hierarchies found in achievement test responses and the sequence and order in observations of child growth. Kingma and Van den Boss (1988) discussed the advantages of the one-parameter logistic to measure growth sequences, and Wohlwill (1960, 1973) successfully applied Guttman scalogram analysis to children's

concept development. Bock & McCabe (1990) indicated that item response theory is an effective foundation for analyzing scores based on authentic student writing samples.

Research Plan

In this study, we address the problem of organizing and analyzing performance assessments by presenting a process that includes (a) a longitudinal data base of performance assessments, (b) an organizational strategy that relies on computers and relational data bases to facilitate analyses of performance assessments, and (c) an application of the Rasch measurement model to identify an underlying growth structure in performance assessments of young children's language behavior. In this process child information is stored on an IBM mainframe computer and manipulated with DB2 software. School personnel or researchers select child information on the mainframe that is transferred to table top computers in local schools. The table top computers in turn have been equipped with Paradox data base software (Borland, 1985-92) and a Rasch measurement computer program, Bigsteps (Linacre, 1992), to identify objective developmental landmarks in the longitudinal record of performance assessments.

The outcome of this process is an objective topography that describes early verbal learning. This structure should be a useful foundation on which to base the evaluation of child growth and plan effective school instruction.

Method

Population

The population are annual cohorts of urban disadvantaged children who enrolled in the following early childhood programs (N = 22,000). Approximately 50% are African American, 38% are Hispanic, and 12% white or other minorities. (These intervention programs and their criteria for enrollment are described in Appendix A.)

- Illinois State Prekindergarten
- Child-Parent Centers
- Head Start
- Chicago Effective Schools Project
- Chapter 1-funded kindergarten[1]
- Chapter 2-funded kindergarten

Data Sources and Collection

Figure 1 presents the information that is collected of the children enrolled in the programs presented above. The following sections provide elaboration.

Field observations. The Bureau of Program Evaluation currently collects the following field observations:

- Observations of Child-Parent Centers and elementary school classrooms
- Parent interviews

Periodic observations are conducted of all early childhood intervention classrooms and a structured interview is conducted with all parents when children enroll in a program.

Standardized tests[2]. For the purposes of program evaluation, the Chicago Board of Education administers the following standardized tests to a sample of children in intervention programs.

- Peabody Picture Vocabulary Test (PPVT-R; Dunn & Dunn, 1981)
- Iowa Tests of Basic Skills (ITBS; Hieronymus, Lindquist, & Hoover, 1980).
- Morris's Ten-Word Instrument (MORRIS; see Morris & Perney, 1984)

Teacher assessment of child performance. Children's competency at performing specific school-related skills are rated by teachers using the following assessment protocols.

---

[1]Chapter 1 and 2 are administrative categories established by the U.S. Department of Education to implement the Educational Consolidation and Improvement Act.

[2]The PPVT-R and ITBS are nationally standardized achievement tests with alpha reliabilities between .85 and .95. The MORRIS is a spelling test of children's phoneme awareness and, in particular, vowel segments based on a standard word list. The MORRIS is not norm-referenced and has shown test-retest correlations that were greater than .80 (Morris & Perney, 1984).

- Chicago EARLY (EARLY; see Chicago Public Schools, 1981)
- Child Development Record (CDR).

Performance samples. The following performance samples are collected from children.

- Composition writing
- Colored pencil drawings
- Clay models and block designs
- Quantitative and analytical thinking

Performance samples represent several domains of children's intellectual development. Writing samples represent children's language development, colored pencil drawings in young children represent conceptual maturity (see Harris, 1963), and clay models, as well as block designs, are used to infer spatial ability. A method of representing quantitative and analytical ability in performance samples needs yet to be developed.

Figure 1

Structure of the Early Childhood Intervention Data Base

Insert Figure 1

General surveys. The following surveys are routinely conducted.

- Child Assessment Profile (CAP; Chicago Public Schools, 1989)
- Parent, teacher, and assistant teacher questionnaires
- Surveys of children's motivation for school and learning

School records. The central computer records of the Chicago Public Schools are regularly searched for the following information.

- Mobility
- Retention
- Special education placement
- Attendance information
- Federally subsidized school lunch status
- Health status

Figure 2 shows a schematic plan for annual collection of performance assessments and other observations.

Figure 2

Schematic Plan of Annual Data Collection

Insert Figure 2

Procedures

Population sampling. All observations and background information annually are entered into an IBM relational data base program (DB2[3]) for mainframe computers. Using IBM DB2 software, representative samples are drawn from the respective programs and transferred to tabletop micro-computers where Paradox, also a relational data base, is used to prepare data for scaling, statistical analyses, as well as horizontal and vertical linking of performance assessments. Figure 3 shows the plan for file transfer between mainframe to tabletop computers.

Figure 3

File Transfer Between Mainframe and Tabletop Computer Data Bases

---

Insert Figure 3

Scale construction and vertical linking. A key procedure is establishing the continuity of behaviors and skills that are common across prekindergarten, kindergarten, and Grade 1. These characteristics should represent an ordered conception of how children progress from low levels to higher levels of development. This procedure statistically analyzes observations in the respective levels to identify latent learning structures. Then it systematically compares the structures to establish continuity and stability. Through a formal linking procedure, an analytical framework is established that is quantitatively uniform across the school levels. Appendix B summarizes the steps followed to identify and link latent learning structures. Figure 4 shows the vertical linking structure that integrates the observations in the school levels into a longitudinal growth continuum. After data is transferred to tabletop computers,

[3]DB2 is commercial software developed by IBM to manage mainframe data bases.

Figure 4

Vertical Linking Structure of the Longitudinal Data Base

Insert Figure 4

---

empirical operations are implemented in the following three phases.

Phase 1. In Phase 1 the Rasch model defines a performance variable based on teacher ratings of cognitive performance collected at the end of prekindergarten. Vocabulary measures are also collected in prekindergarten to establish a common performance variable and to identify prospective anchor items. Because both ratings and test scores measure aspects of language ability, some of their items should empirically define a common variable.

Phase 2. Phase 2 consists of analyzing writing, spelling, and drawing samples collected in kindergarten to establish a common variable. Rasch item difficulties of the anchor items are compared between Phase 1 and 2. The result of Phase 2 is a continuous dimension of verbal development between prekindergarten and kindergarten.

Phase 3. Phase 3 focuses on assessments in grade 1, a dimension again defined by writing, spelling, and drawing samples, which is equated with the Phase 2 performance variable. The result is a common structure of language development that extends from prekindergarten (age 3) through grade 1.

All scaling and linking is conducted with Bigsteps (Linacre, 1992).

Analyses

The following analyses are conducted on the data base.

Phase 1. The prekindergarten teacher assessment items and PPVT-R scores are calibrated independently of each other. Item calibrations of the two forms are plotted and items are identified that define a common variable. These items are calibrated as a group, and the obtained person measures establish positions on a Phase 1 language performance variable.

Phase 2. Teacher assessments collected during kindergarten are calibrated and common items between Phases 1 and 2 are examined for stability and item calibrations are correspondingly equated. Next characteristics that are coded in the performance samples are calibrated as a group with the teacher assessment items and the obtained person measures establish positions on a continuous performance variable between Phases 1 and 2.

Phase 3. The characteristics that are coded in grade 1 are calibrated and equated with Phase two characteristics. The obtained person measures establish

positions on a continuous performance variable from Phase 1 to Phase 3.

Results

Table 1 presents overall descriptive statistics of the writing, drawing, and teacher assessments in the data base for a sample of kindergartens.

Table 1

Descriptive Statistics

Table 1

Figure 5 shows the results of a pilot study that examines the fit of 35 characteristics in kindergarten writing samples, colored drawings, and teacher survey responses to the Rasch model. In general they tend to define a developmental progress of language development from low to high. At the bottom of the scale or less-mature level of development are attitudes and behaviors that are associated with readiness for school learning (i.e., identies basic shapes, eager to learn, works diligently, and so on). Higher up in the structure appear language-related skills such as uses invented spelling and liklihood of learning to read. Moreover, in this type of analysis, children who are positioned at any particular point in the structure tend to show the lower characteristics more frequently than characteristics above their position.

Figure 6 shows a comparison of characteristics that are common to assessments in preschool and kindergarten in bicalibration plots. The results indicate that, despite differences in time, the sequence of the characteristics remain essentially invariate. This means that as children's verbal development becomes more sophisticated they tend not to lose their ability to perform lower level skills.

Figure 5

Performance Variable Empirically Defined by Drawing, Writing, and Spelling Samples

---

Insert Figure 5

Figure 6

Bicalibration Plots Between Common Items in Preschool, Kindergarten, and Grade 1 Performance Samples

---

Insert Figure 6

Figure 7 shows an hypothesized structure of language development that we expect to obtain when assessments are linked across school levels. Not surprisingly, we expect to find a structure typical of learning curves (i.e., rapid growth early in a cycle gradually reaching a plateau).

Figure 7

Language Development Across Preschool, Kindergarten, and Grade 1

Insert Figure 7

Figure 8

Sample Teacher-Parent Form

Insert Figure 8

Discussion

DRAFT

This study has presented a process (i.e., performance assessments scored and centrally stored in a large computer file that are examined in meaningful units of analysis by the Rasch measurement model) that integrates information collected from disparate observations and surveys of early childhood intervention programs into a coherent structure of early school verbal development. The process is surprisingly innovative. A review of literature revealed many medical and psychiatric applications of relational data bases to facilitate diagnoses and treatments but few studies of assessment or child development data bases (Endo & Bittner, 1985; Winer & Carriere, 1990), and none integrating assessments across school or developmental levels. Despite an application of statistical measurement technology, this process not only supports the philosophy of authentic assessment but actually addresses the following challenge presented by evaluation reformers:

> To document [student] accomplishment, we must also design assessments that are *longitudinal* (italics added), sampling the baseline, the increment, and the preserved levels of change that follow from instruction . . . monitoring whether children have passed the milestones that would allow them to change the level of their current reading and writing activities (Wolf, Bixby, Glenn, and Gardner, 1991, pp. 52-53).

The results of this process provide teachers and parents with information that promote an integration of assessment into the planning of school learning and presumably developmental enhancement.

Because total enrollment in the Chicago Public Schools is over 400,000 students in over 300 elementary schools, we have relied on a mainframe computer for central storage of assessment information. We fully expect in the future that microcomputer networks will have the storage capability necessary to conduct this process. A large centralized mainframe computers archives child assessment records and maintains confidentiality of sensitive child performance information. Because DB2, the mainframe data base software, only provides child information that is selected by researchers or teachers at a particular time yet can provide additional information whenever necessary it provides flexibility and efficiency. Because external background information, as well as a broad array of developmental variables can be easily stored, this process provides researchers with an opportunity to speculate on longitudinal relations among variables that may not become important for many years. The implementation of table top computers (1) facilitates the analysis of variables, (2) offers schools an opportunity to share information, and (3) dramatically reduces the cost of computer operations. Finally, because mainframe storage is centralized, contemporary data files are compatible with historical data files, and thus child background variables such as mobility, welfare status, shifts in socioeconomic status, referral for special services (i.e., counseling, special education, and so on) can be readily examined.

General Issues

A process has been described to identify a structure of school-related verbal learning in the performance assessments and teacher surveys of young children. Using modern measurement technology, a sequence of language-related skills was identified and quantitatively linked into a precise and rigorous analytical framework. This hierarchy is developmentally meaningful -- a plausible sequence of growth, empirically based, and dependent on school instruction. This approach should prove instrumental in providing continuity and clarity to both the interpretation of child growth and the evaluation of intervention programs.

The results are theoretically interesting because they suggest that the emerging language of children, represented by their authentic performances, show patterns of continuity across annual assessments and thus identify a fundamental trajectory of verbal development.

An important aspect of this process is the considerable promise it offers as a heuristic device to evaluators trying to establish the specific performances that need to be included in an assessment design. Consequently, this process should increase the validity of evaluations. Another benefit of this process is its capacity to identify redundancy or functionally equivalent aspects of performance assessments (Miller-Jones 1989; see Linn, Baker, Dunbar, 1991) thus promote the refinement and economy of the assessment process.

These results, however, raise the following questions:

1. Should assessments focus on developmental landmarks unaffected by environment or only on instructional skills that are influenced by school learning?

2. How should this process be implemented to identify multiple growth trajectories (i.e, quantitative, spatial, and analytical abilities)?

3. Does this process address concerns about the stability of performance assessments?

References

Biddle, B.J., & Marlin, M.M. (1987). Causality, confirmation, credulity and structural equation modeling. *Child Development, 58*, 4-17.

Bock, R.D., & McCabe, P. (1990). *Toward a timely state NAEP.* Unpublished manuscript prepared for the National Assessment Governing Board.

Borland, Inc. (1985-92). *Paradox* Version 4.0. San Francisco: Borland International, Inc.

Burchinal, M., & Appelbaum, M.I. (1991). Estimating individual developmental functions: Methods and their assumptions. *Child Development*, 62, 23-43.

Chicago Board of Education (1981). *EARLY Assessment and Remediation Laboratory.* Chicago: The Board.

Chicago Board of Education (1989). *Child Assessment Record.* Chicago: The Board.

Dunn, L.M., & Dunn, L.M., (1981). *Peabody Picture Vocabulary Test Revised.* Circle Pines, Minnesota: American Guidance Service.

Endo, J., & Bittner, T. (1985). Developing and using a longitudinal student outcomes data file: The University of Colorado experience. In P.T. Ewell (Ed.), *New Directions for Institutional Research.* San Francisco: Jossey-Bass.

Harris, D.B. (1963). *Children's drawings as measures of intellectual maturity. A revision and extension of the Goodenough Draw-a-Man Test.* New York: Harcourt, Brace, & World.

Herman, J.L. (1992). What research tells us about good assessment. *Educational Leadership, 49*, 74-78.

Hieronymus, A.N., Lindquist, E.F., & Hoover, H.D. (1980). *Iowa Test of Basic Skills: Primary battery.* Chicago: Riverside.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology, 27*, 236-248.

Kingma, J., & Van Den Bos, K. (1988). Unidimensional scales: New methods to analyze the sequences in concept development. *Genetic, Social, and General Psychology Monographs, 114*, 479-509.

Linacre, J.M. (1992). *Bigsteps* Version 2.27. Chicago: MESA Press.

Linn, R.L., Baker, E.L., & Dunbar, S.B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20*, 15-21.

Loevinger, J. (1968). The meaning and measurement of ego development. *American Psychologist, 21*. 195-206.

Martin, J.A. (1987). Structural equation modeling: A guide for the perplexed. *Child Development, 58*, 33-37.

Mehrens, W.A. (1992). Using performance assessments for accountability purposes. *Educational Measurement, 11*, 3-9.

Miller-Jones, D. (1989). Culture and testing. *American Psychologist*, 44, 360-366.

Morris, D., & Perney, J. (1984). Developmental spelling as a predictor of first-grade reading achievement. *Elementary School Journal, 84*, 441- 457.

Moss, P.A., Beck, J.S., Ebbs, C., Matson, B., Muchmore, J., Steele, D., Taylor,C., & Herter, R. (1992). Portfolios, accountability, and an interpretive approach to validity. *Educational Measurement, 11*, 12-21.

O'Neil, J. (1992). Putting performance assessment to the test. *Educational Leadership, 49*, 14-19,

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests.* Chicago: University of Chicago Press.

Rutter, M. (1988). *Psychosocial risk: The power of longitudinal data.* Cambridge, England: Cambridge University Press.

Von Eye, Alexander (1990). (Ed.). *Statistical methods in longitudinal research.* Orlando, FL: Academic Press.

Winer, L.R., & Carrier, M. (1990). The use of a relational database in qualitative research on educational computing. *Computers in Education, 15*, 213-220.

Wohlwill, J.F. (1960). A study of the development of the number concept by scalogram analysis. *Journal of Genetic Psychology, 97*, 345-377.

Wohlwill, J.F. (1973). *The study of behavioral development.* New York: Academic Press.

Wolf, D., Bixby, J., Glenn, J., & Gardner, H. (1991). To use their minds well: Investigating new forms of student assessment. In G. Grant (Ed.) *Review of Research in Education, 17*, 31-74.

Wright, B.D., & Stone, M. (1979). *Best test design.* Chicago: MESA Press

Table 1

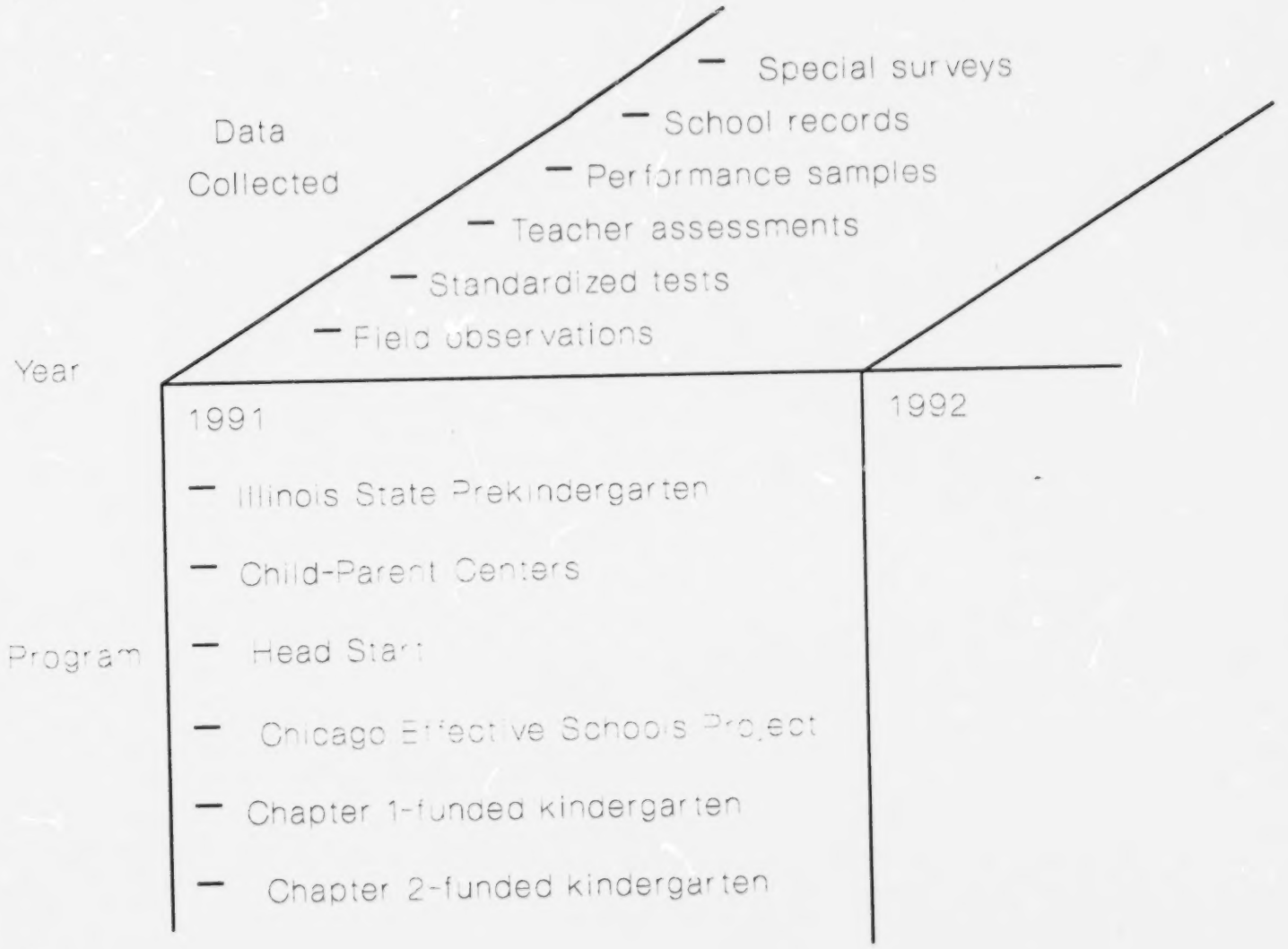
Descriptive Statistics

Data sources	Mean	SD	Rasch Calib	Fit value	N
State Prekindergarten					
Teacher assessments	.XX	.XX	.XX	.XX	X
Identifies basic shapes					
Recognizes numbers					
Eager to learn					
Shows self-confidence					
Colored pencil drawings					
Shows activity center					
Proportional relations					
symbolic content					
Human figure detail					
PPVT-R					
Kindergarten					
Teacher assessments		--- To be included--			
Liklihood child will read					
Identifies basic shapes					
Recognizes numbers					
Eager to learn					
Shows self-confidence					
Colored pencil drawings					
Shows activity center					
Proportional relations					
Symbolic content					
Human figure detail					
Writing samples					
Awareness of print					
Directionality					
Content					
PPVT-R	.XX	.XX	.XX	.XX	X

Grade 1
- Teacher assessments
  - Liklihood child will read
- Colored pencil drawings
  - Shows activity center
  - Proportional relations
  - Symbolic content
  - Human figure detail
- Writing samples
  - Awareness of print
  - Directionality
  - Content
- ITBS

Note. Abbreviations: ITBS (Iowa Tests of Basic Skills), PPVT-R (Peabody Picture Vocabulary Test-Revised).

Figure 1: *Structure of the Early Childhood Data Base*



Note. *N*s range from 164 (Chapter 2-funded kindergarten) to 10,000 (Illinois State Prekindergarten).

Figure 2: *Schematic Plan of Annual Data Collection*

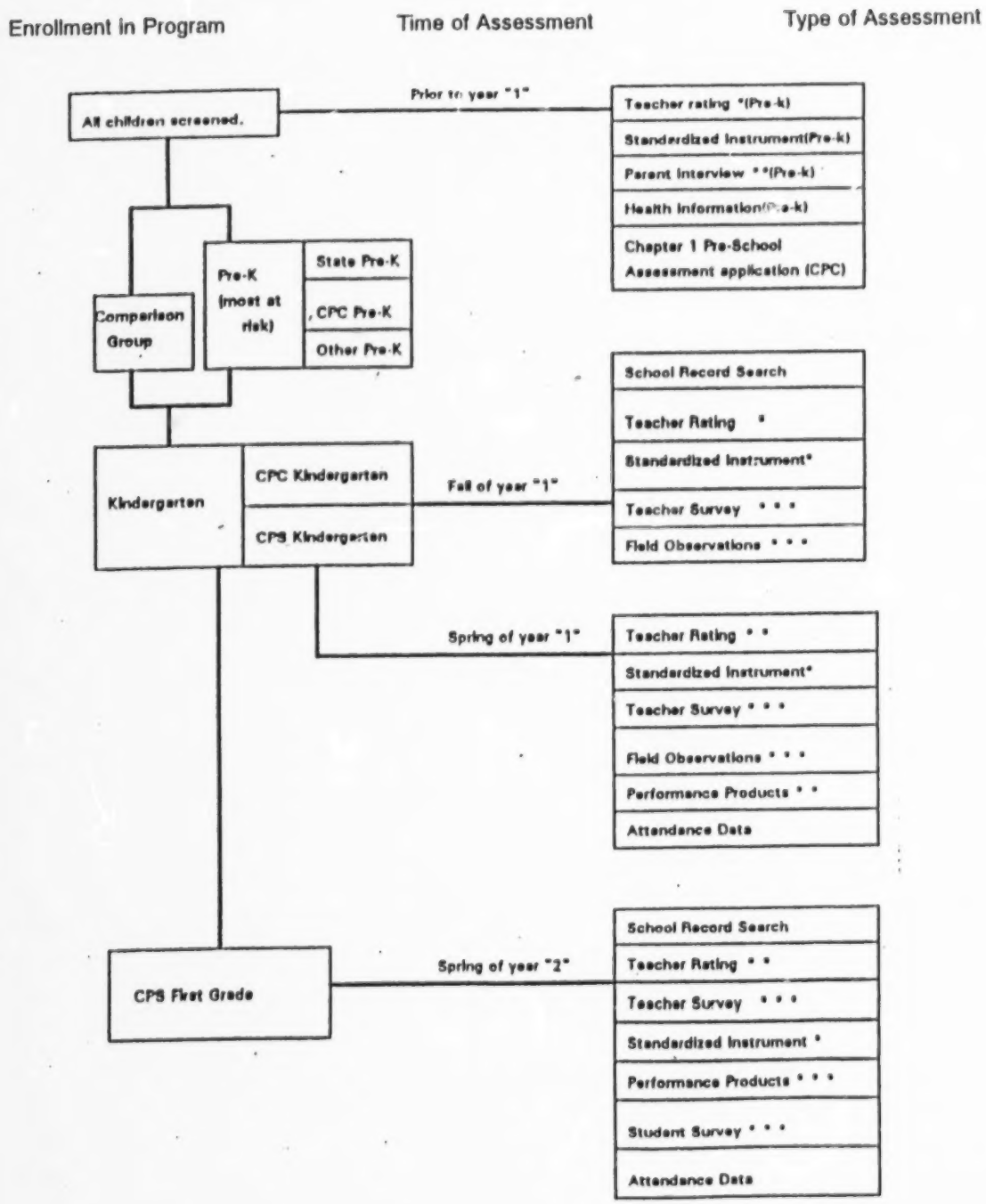
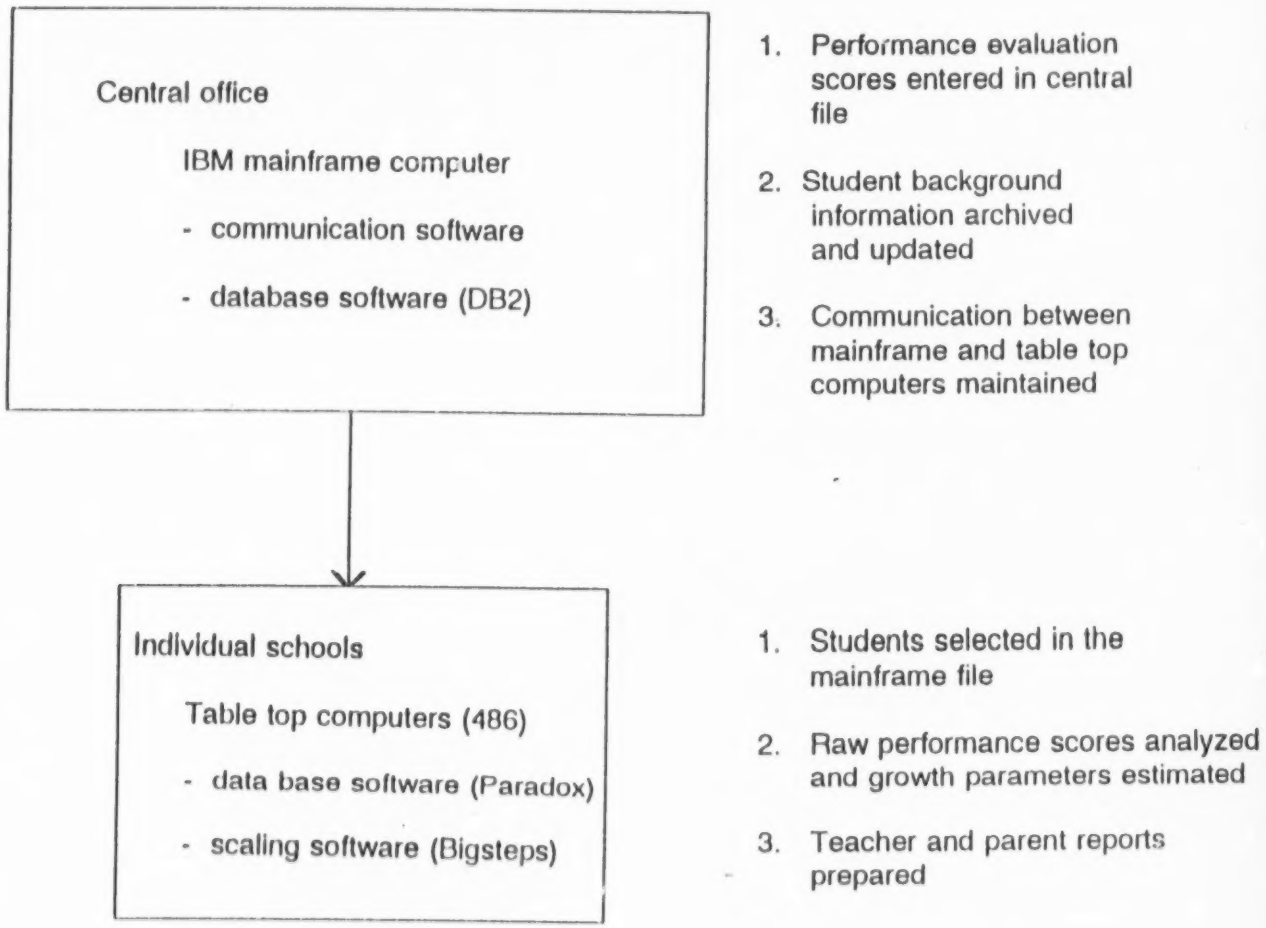
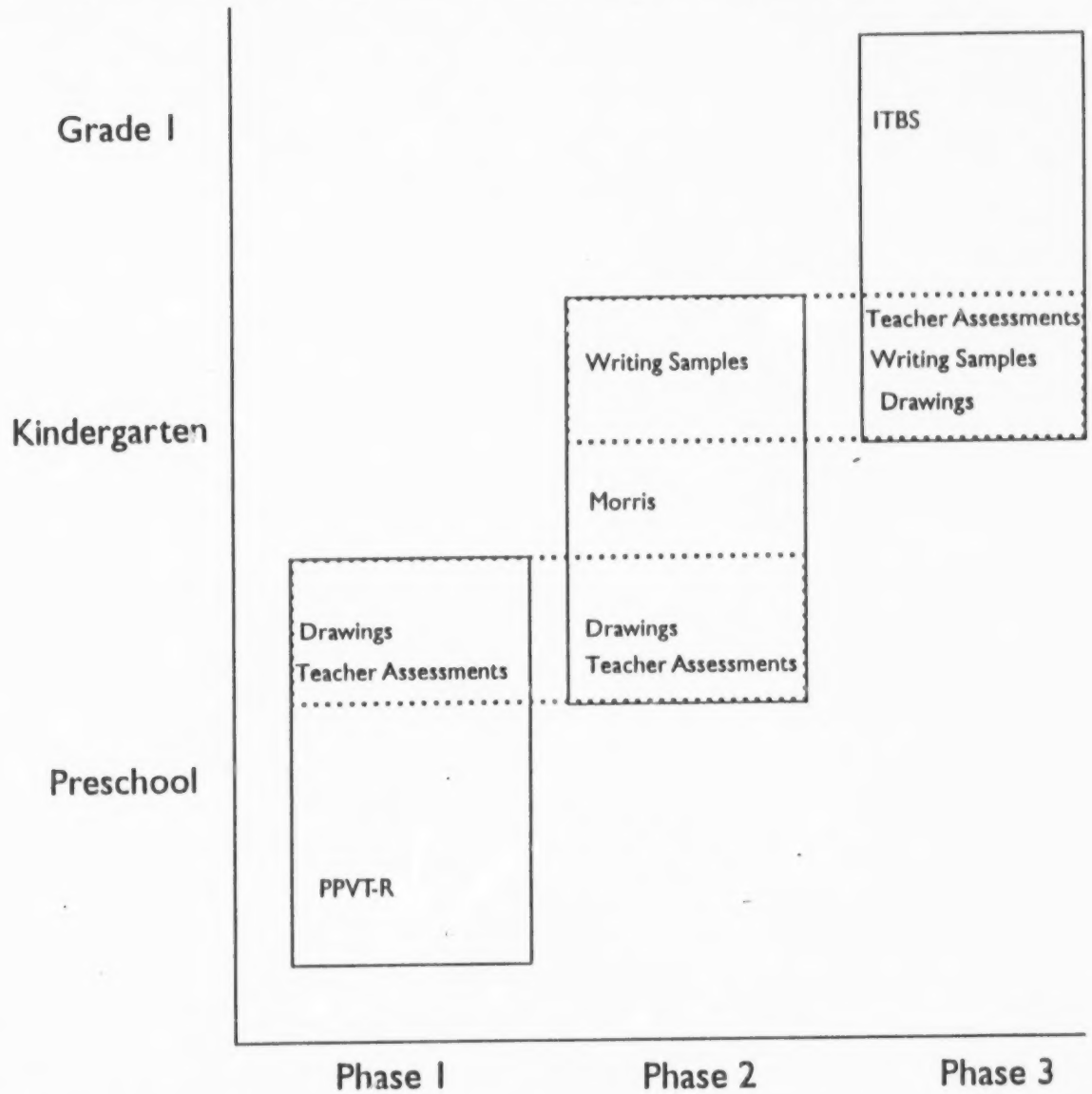


Figure 3: *File Transfer Between Mainframe and Tabletop Computer Data Bases*


Central office
IBM mainframe computer
- communication software
- database software (DB2)
1. Performance evaluation scores entered in central file
2. Student background information archived and updated
3. Communication between mainframe and table top computers maintained
Individual schools
Table top computers (486)
- data base software (Paradox)
- scaling software (Bigsteps)
1. Students selected in the mainframe file
2. Raw performance scores analyzed and growth parameters estimated
3. Teacher and parent reports prepared

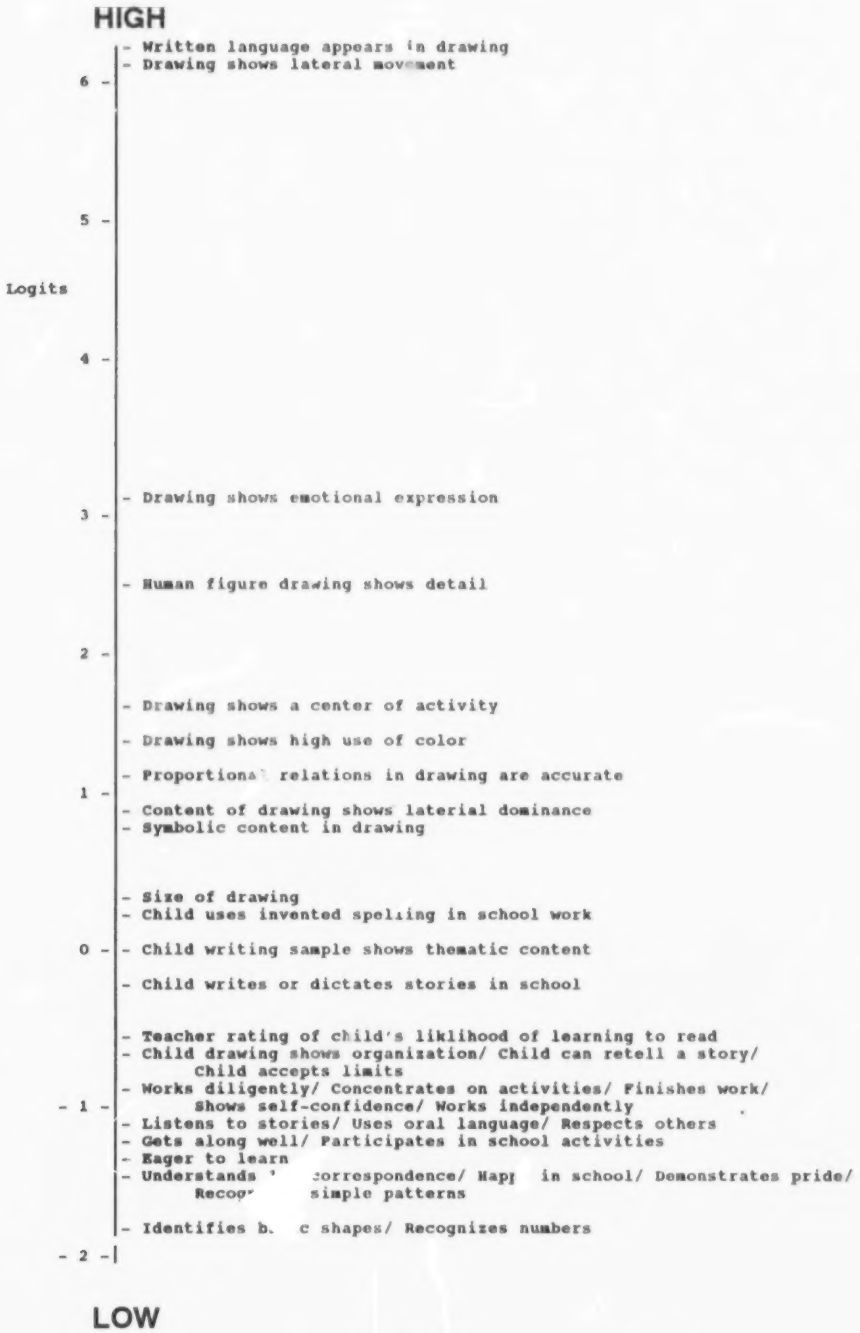
Note. Over 300 public elementary schools (N > 410,000) are currently linked to a central office IBM mainframe computer. DB2, Paradox, and Bigsteps are commercially available computer software.

Figure 4: *Vertical Linking Structure of the Longitudinal Data Base*



Note. The phases represent independent scaling of preschool, kindergarten, and grade 1. Abreviations: ITBS (Iowa Tests of Basic Skills), PPVT-R (Peabody Picture Vocabulary Test-Revised), Morris (Morris Basic Ten Words).

Figure 5: Kindergarten Performance Variable Defined by Drawing Sample, Writing Sample, and Teacher Assessment


HIGH
- Written language appears in drawing
- Drawing shows lateral movement
6 -
5 -
Logits
4 -
- Drawing shows emotional expression
3 -
- Human figure drawing shows detail
2 -
- Drawing shows a center of activity
- Drawing shows high use of color
- Proportional relations in drawing are accurate
1 -
- Content of drawing shows laterial dominance
- Symbolic content in drawing
- Size of drawing
- Child uses invented spelling in school work
0 - - Child writing sample shows thematic content
- Child writes or dictates stories in school
- Teacher rating of child's liklihood of learning to read
- Child drawing shows organization/ Child can retell a story/
Child accepts limits
- Works diligently/ Concentrates on activities/ Finishes work/
- 1 - Shows self-confidence/ Works independently
- Listens to stories/ Uses oral language/ Respects others
- Gets along well/ Participates in school activities
- Eager to learn
- Understands ' correspondence/ Happ in school/ Demonstrates pride/
Recog simple patterns
- Identifies b. c shapes/ Recognizes numbers
- 2 -
LOW

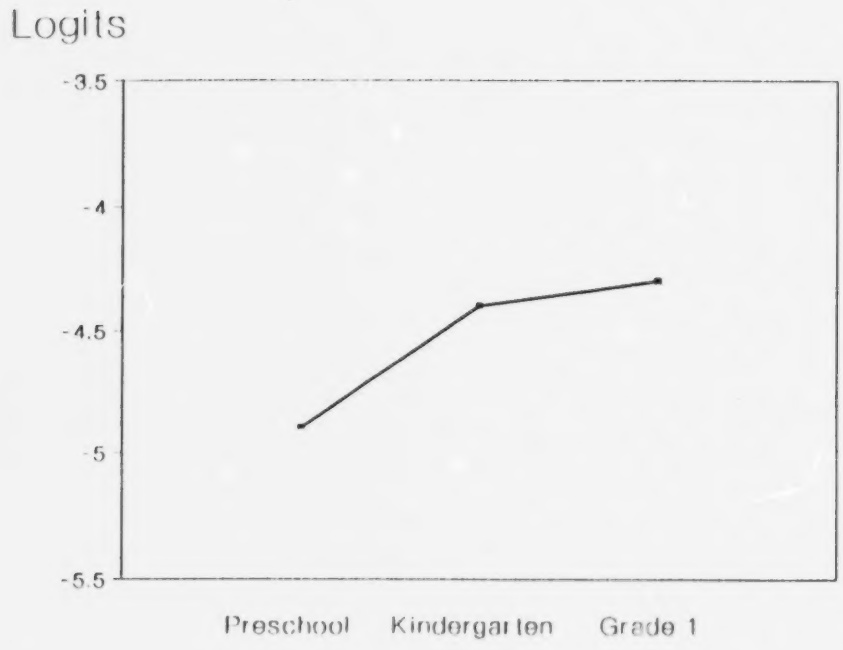
Note. $N = 202$. Zero is the mean difficulty of 35 items calibrated on the kindergarten group. All values are writing, drawing, and spelling performance assessments as well as teacher ratings that were transformed to one-parameter logits. Average Rasch person infit value = -.2, $SD = 1.9$. (*Mnsq* = 1.01, $SD = .49$), and average item infit value = -.7, $SD = 4.5$ (*Mnsq* = 1.04, $SD = .53$). Reliability of person and item separation are .95 and 1.00, respectively. Alpha reliability of 35 items is .96.

Figure 6: *Bicalibration Plots Between Common Items in Preschool, Kindergarten, and Grade 1 Performance Samples*

To be included

Note.

Figure 7: *Language Development Across Preschool, Kindergarten, and Grade 1*



Note. This plot represents a hypothesized growth structure based on assessments of drawing, writing, and spelling performance samples that are scored and transformed to one-parameter logits. In this approach to modeling child growth the Rasch model identifies an objective verbal language structure among the performance measures and transformed teacher ratings and longitudinally links them across preschool, kindergarten, and grade 1. Zero represents an hypothesized elementary school mean.

Figure 8: *Sample Parent Form*

ANY PUBLIC SCHOOL

Department of Instruction and Assessment

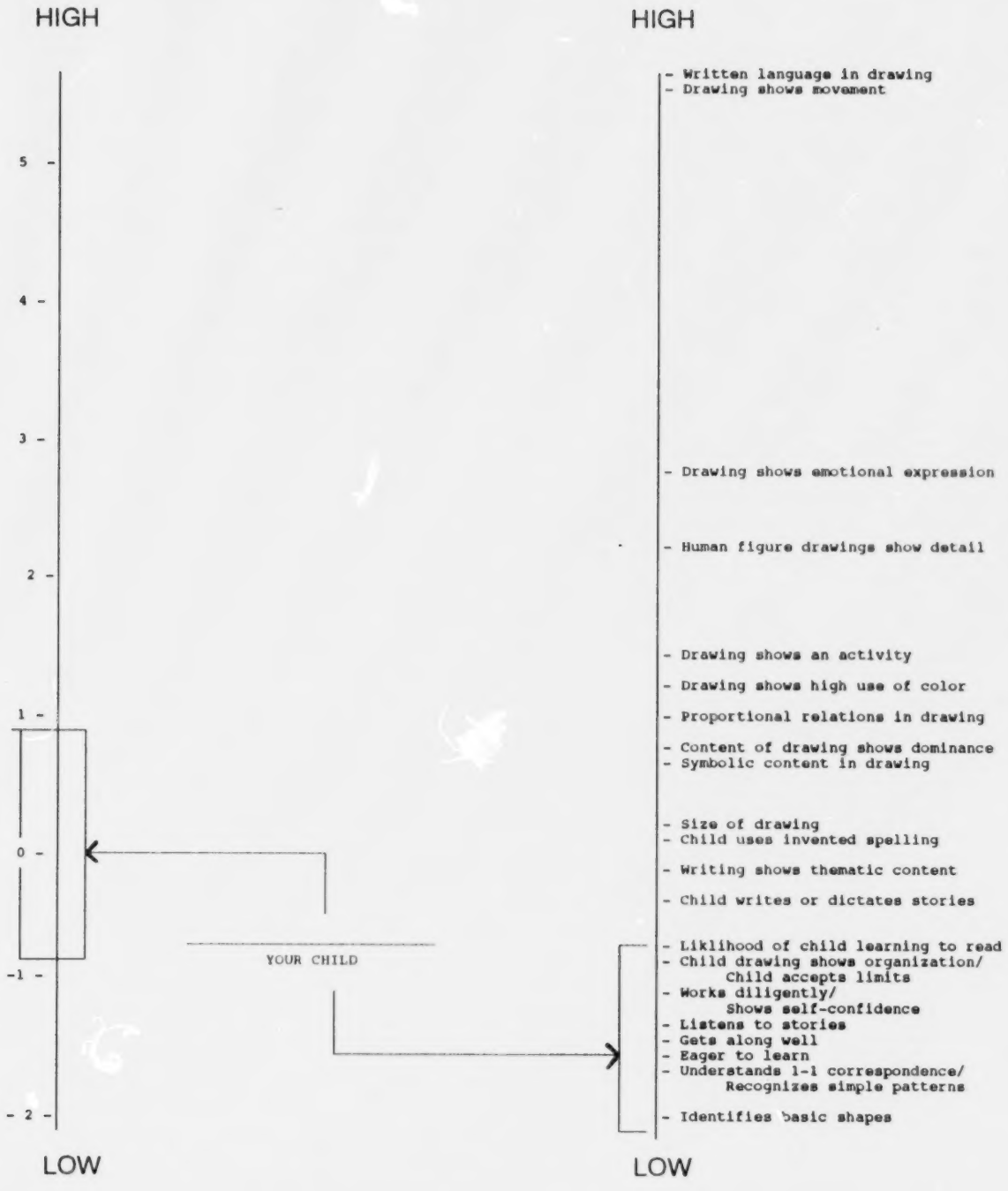
Parent Report Form

To the parents of _____.

Instructions: The vertical scales on the following page represent language development and school learning. Note your child's position on them. The vertical scale on the right shows a sequence of school learning. The boxed area indicates average learning for your child's school level. The skills and behaviors below your child's marked position have been observed in his or her school performances. They tend to appear more frequently than the skills and behaviors above his or her position.

The scale on the left represents the results of a nationally normed test of language development that is *not* based on school experiences. The boxed area represents average language development. Study the position of your child relative to the scales and arrange an appointment to meet with his or her teacher. Be prepared to discuss your interpretation of your child's progress represented in this report. Provide enough time to discuss the home learning experiences that may be appropriate to enhance your child's school learning and developmental growth.

Figure 8: *Sample Parent Form* (continued)



Appendix A: *Description of Early Childhood Intervention Programs*

The Chicago Public Schools Department of Early Childhood currently enrolls eligible children in the following early intervention programs.

Illinois State Prekindergarten (State Prek). Approximately 10,000 children in Chicago are enrolled annually in State of Illinois-funded preschools. Although the program has no income requirements, children are screened to establish enrollment eligibility.

The purpose of the program is to identify and prepare educationally at-risk three- to five-year old children for school success. Program features include experiences that promote children's positive self-image and learning in language and communication skills.

Socioeconomically the population is 50% African-American, 36% Hispanic, 11% Nonminority-American. Economically, 80% of the families are eligible for 100% subsidized school lunch and 7% are eligible for a reduced-price school lunch.

Child-Parent Centers (CPCs). In 1991, 2,939 preschoolers and 875 kindergartners were enrolled in federally funded Chapter 1 CPCs, and 357 were enrolled in Chapter 2 CPCs. Any family that resides in a federally designated poverty neighborhood is eligible for enrollment, and the population is predominantly African-American (>.90). At least 50%, and sometimes as many as 90% of the children were living in public housing projects when they enrolled in kindergarten.

The goal of this program is to improve children's cognitive, as well as affective, readiness for the primary grades through an educational emphasis on language and basic skills. The program provides teachers with assistants, support services, and inservice meetings conducted by the Bureau of Early Childhood Programs of the Chicago Public Schools. All CPC centers have a parent-involvement component.

Head Start (HS). Approximately 3,000 students are enrolled in the federally funded Head Start program. Eligibility is based entirely on economic need as established by the U.S. Department of Human Services. Unlike the other programs, 10% of the children have been recruited because they are handicapped by a physical or mental disability. Only children with the greatest economic need are enrolled, and more than 95% of Head Start families are receiving public assistance.

The program provides 2-1/2 hours of preschool experience, has a parent component, as well as medical, dental, and psychological services.

In addition to classroom staff consisting of certificated teachers and assistants, a parent development team, social worker and assistant provide services.

All parents are required to participate as volunteers in classrooms and monthly parent meetings are led by the parent development team.

Chicago Effective Schools Project (CESP). CESP represents schools that have been selected for special federal desegregation funds because their ITBS scores are the lowest in the city. Current enrollment is 1,710 children and the program consists of an all-day kindergarten using the conventional CPS curriculum. None of the CESP schools has a parent-involvement component.

Chapter 1-funded kindergarten. Chapter 1 federal funds are used in kindergartens

for 2,233 children. Funds supplement school budgets, although specific expenditures for classroom services are not specified.

Chapter 2-funded kindergarten. In 1991, Chapter 2 federal funds were also used in kindergartens for 164 children. This program differs substantially because preschool is not offered and the program is open to all families regardless of financial status. Chapter 2 kindergarten (no preschool) primarily emphasizes teacher inservices.

Appendix B: Description of the Linking Procedure

The linking procedure is based on a mathematical approach to item scaling originally developed by Rasch (1960/1980) to solve the problem of measuring mental ability without referring to specific test items or norm groups. His solution was to compare patterns of consistency between right and wrong answers to test items with values predicted by the one-parameter logistic function below:

BTD, p.15 -- insert equation here --

In his approach, the probability of passing or missing an item is mathematically modeled as the simple difference between a person's ability and an item's difficulty. This formulation, unlike others such as percents or total scores, transforms test scores which are restricted between 0 and 100 percent and thus nonlinear into measures which run from minus infinity to plus infinity. This representation of the interactions between persons and items makes measurement specificly objective.

Because the difficulty of items and in this study the characteristics of performance samples differ depending on school and developmental level of children, item difficulties need to be equated before the ability of groups can be compared. This is formally accomplished through the following linking procedure:

BTD, p. 96 -- insert equation here --

By using anchor items (identical items administered to different groups), this procedure establishes a common metric for different children across levels of development. By conceptualizing characteristics in children's performances such as writing or drawing as dichotomous events, this approach provides a model for comparing any child to an objective structure of development established by his or her cohort.

By approaching children's performance samples as the appearance or nonappearance of particular characteristics, statistical parameters can be estimated for the abilities necessary to produce them. When performance samples share some common characteristics, they may be mapped into a common analytical framework that can be linked across groups of children to identify developmental levels.

Author Note

We gratefully acknowledge the support and cooperation of the Chicago Public Schools in conducting this study. In particular, Jeanne Borger, Coordinator, Department of Research, Evaluation and Planning provided many helpful comments. Nancy Mavrogenes and Marcia Kurland also provided valuable assistance concerning the description of early childhood programs in the Chicago Public Schools.

Further information concerning this study are available from the first author at 1532 E. 59th Street, Chicago, Illinois 60637.